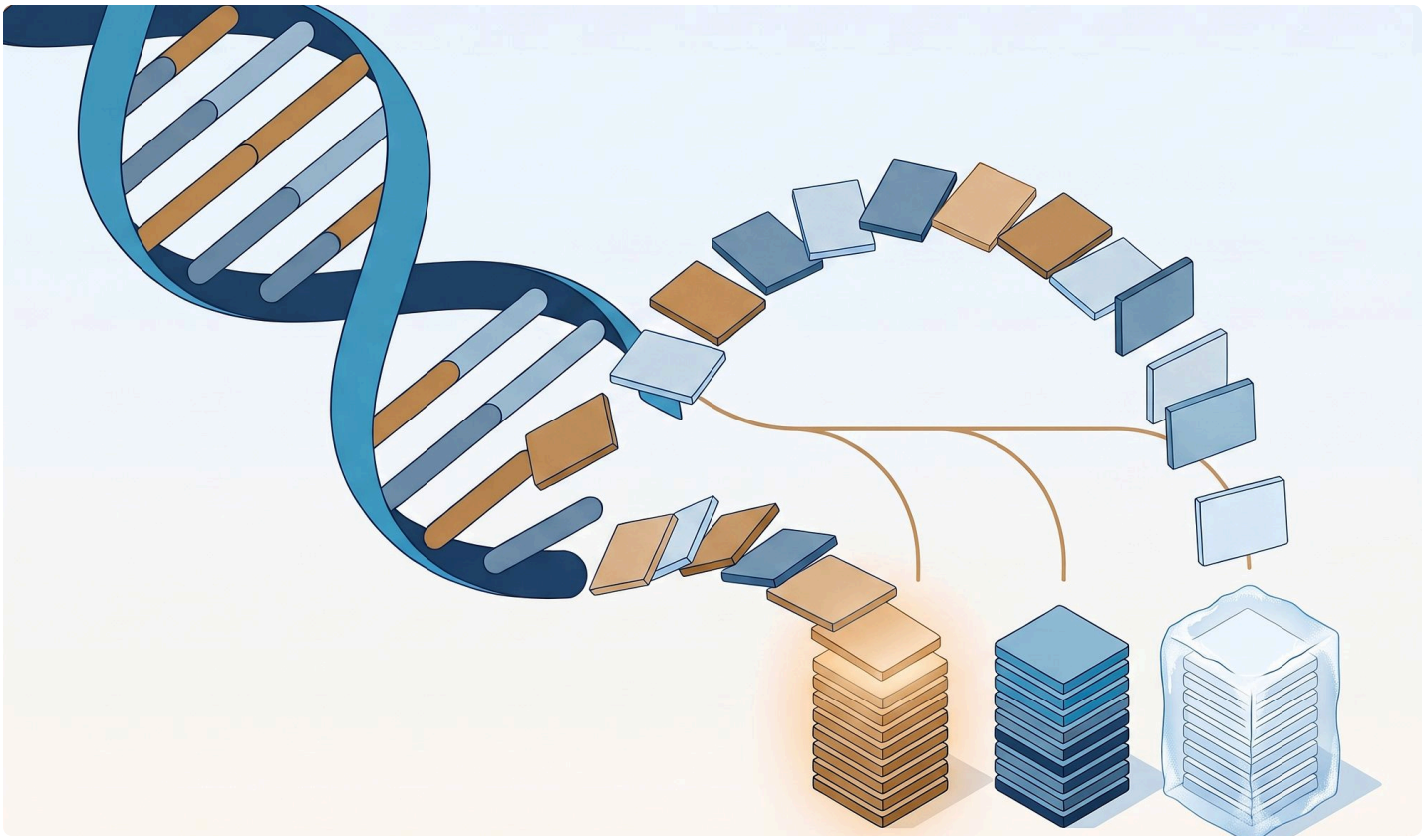
Case Study: The Salk Institute and Starfish

# Archiving and Tiered Storage Management for Genomics Research



## Client

The Salk Institute for Biological Studies - Integrative Genomics and Bioinformatics Core

## Challenge

Rapid genomics data growth overwhelming expensive high-performance storage and filling archive storage with redundant, unneeded BCL files.

## Solution

Starfish used to automate data cleanup, archival to Amazon Deep Glacier, and ongoing cataloging of genomics datasets.

## Results

Reclaimed ~200 TB of storage; automated data management processes; gained usage insights for accurate capacity planning.


© 2026 Starfish Storage Corporation. All Rights Reserved.

# The Challenge

April Williams Wehner, PhD, Core Director for Integrative Genomics and Bioinformatics at the Salk Institute for Biological Studies, faced a critical and costly data crisis. For over a decade, the Institute's simple, effective storage strategy relied on continually purchasing new, faster storage systems for active data and reassigning the old, filled-up systems to serve as permanent on-premises archives.

However, rapid advancements in sequencing technology, with instruments capable of generating up to **10 terabytes (TB) of raw data per experiment**, made this model ineffective. What was once a five-year investment in 500 TB of active storage was now being filled in just two years. The Institute was on a trajectory to spend millions of dollars simply to keep pace with the growth of rarely accessed, yet scientifically critical data.

This acute storage capacity problem was compounded by years of data management debt: the Institute had never deleted any of the raw, binary BCL genomic sequencing files, even after they were converted into FastQ files for analysis. This debt resulted in the retention of two copies of the same information in different formats, one usable and one not, cluttering both archive and active storage, and adding complexity and inefficiency to a system that needed simplicity to scale.

> "Preserving sequencing data is important because experiments can be costly to repeat. A single one of the large sequencing experiments can cost $25,000. If we lose that data, then that's a big loss for us."
> — **April Williams Wehner, PhD**, Core Director for Integrative Genomics and Bioinformatics, Salk Institute



# The Starfish Solution: Intelligent Automation

Williams and her team determined that the retained BCL files were only required for a six-month period for reprocessing in case of errors, and that **keeping the files indefinitely created an unnecessary drain on resources**. Starfish was used to scan and catalog the existing storage systems to identify which files could be safely removed.

Starfish was then used to automate several data workflows. First, as soon as data is processed into usable FastQ files, Starfish automatically uploads an offsite copy to Amazon Deep Glacier, providing reliable protection from catastrophic data loss. Second, after six months, when much of the run data is considered "cold," Starfish automates the movement from "hot" active storage to the older, on-premises archival layer. Third, Starfish automatically deletes BCL files that are more than six months old.

### Amazon Deep Glacier

Automatic offsite backup for every processed sequencing run, protecting against catastrophic data loss.

### On-Premises Archive

Older storage systems repurposed as a cost-effective archival tier for cold but scientifically critical data.

### Starfish Catalog

Ongoing indexing of all "omics" data for centralized management, efficient sharing, and accurate capacity planning.

# The Results

**Reclaiming 200 TB enabled the Institute to defer capital expenditure on new infrastructure**. Additionally, the automated cleanup of BCL files slows the pace of future storage growth, and the team now has the usage data needed for more accurate capacity planning.

Additionally, the solution increased researcher confidence and efficiency. Researchers now know their data is reliably backed up and safe from loss. Biologists and post-docs have more time to focus on research, since they no longer manually archive files or rely on risky personal storage. Despite the movement to archive, **researchers notice no difference in accessing their files**, thanks to a Starfish script that generates soft links, which makes their movement invisible to end-users.

The Starfish data catalog now gives April's research team deep visibility into its "omics" data which enables more centralized management, more efficient data sharing, and reduced dependence on any single individual to know where historical data resides.

In summary, implementing Starfish allowed The Salk Institute to introduce well-thought-out automated processes for managing large research datasets and maintaining archive storage.

ⓘ *Mention of specific commercial products or services does not imply endorsement by the Salk Institute for Biological Studies.*


© 2026 Starfish Storage Corporation. All Rights Reserved.

# About The Salk Institute

The Salk Institute for Biological Studies, located in La Jolla, California, is an independent, non-profit scientific research facility founded in 1960 by Jonas Salk, the developer of the first safe and effective polio vaccine. The institute is dedicated to foundational, collaborative research in critical areas such as aging, cancer, immunology, neuroscience, and plant biology.

# About Starfish

Starfish Storage provides the world's most scalable and versatile file management platform. The platform features an Unstructured Data Catalog (UDC) for indexing and extracting metadata from multi-vendor, multi-PB storage environments, combined with Starfish's powerful Automation Engine for automating data flows. Starfish Zones provide end users with actionable views of their data. Starfish empowers organizations to gain control of their files on a large scale, in mixed vendor environments, optimize storage costs, and enhance data management practices. For more information about how your organization can unlock the power of its data, please visit Starfish Storage at https://starfishstorage.com/.

© 2026 Starfish Storage Corporation. All Rights Reserved.